

Efficiently connecting textual and visual information in operating instructions

Tenbrink, T.; Maas, A.

IEEE Transactions on Professional Communication.

DOI:
[10.1109/TPC.2016.2517451](https://doi.org/10.1109/TPC.2016.2517451)

Published: 29/01/2016

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Tenbrink, T., & Maas, A. (2016). Efficiently connecting textual and visual information in operating instructions. *IEEE Transactions on Professional Communication.*, 58(4), 346 - 366.
<https://doi.org/10.1109/TPC.2016.2517451>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

"(c) 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works."

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**EFFICIENTLY CONNECTING TEXTUAL AND VISUAL INFORMATION IN
OPERATING INSTRUCTIONS**

Thora Tenbrink	Annika Maas
School of Linguistics and English Language Bangor University College Road Bangor, Gwynedd LL57 2DG United Kingdom Telephone: +44 1248 382263 E-mail: t.tenbrink@bangor.ac.uk	FB10 Faculty of Linguistics and Literary Sciences University of Bremen Postbox 330440 28334 Bremen Germany Telephone: +49 421 69313623 E-mail: annika.maas@googlemail.com

Acknowledgements: We thank the company Miele & Cie. KG for their kind support throughout this research, and permission to reprint figures 4, 5, 8, 13 (right), 14, and 18. Figures 1, 2, 3, 6, 7, 9, 10, 13 (left), 15, 16, and 17 were reproduced to match the style of the depictions used in Miele operating instructions.

EFFICIENTLY CONNECTING TEXTUAL AND VISUAL INFORMATION IN OPERATING INSTRUCTIONS

ABSTRACT

Research problem: Most technical documents rely on a combination of text and visuals to communicate their messages. To achieve the desired effect of improved processing and comprehension of operating instructions, the text must guide readers in a clear way to the relevant visual information in order to avoid ambiguity and misinterpretation, and to ensure that the reader optimally benefits from the available information.

Research questions: (1) How are textual and visual information combined in operating instructions so as to guide the user's attention precisely toward the relevant parts of a household appliance? (2) In what ways can ambiguity arise, and what kinds of strategies can be used to avoid miscommunication and conceptual problems?

Literature review: Operating manuals are usually procedural instructions that tell the user how to set up an appliance, how to operate and maintain it, and how to solve any problems. The vast majority of operating instructions are multimodal in that they include visuals of some kind. But previous research has shown that visual representations of instructions can be just as difficult to interpret as textual instructions – especially if the visuals come without textual elaboration. A combination of text and visuals provides opportunities for taking advantage of both by balancing the other mode's limitations.

Methodology: Our exploratory qualitative study of a small set of operating instructions identifies and classifies types of references to visuals in the documents, enriched by interviews with the technical writers of these manuals. Besides showing patterns of reference types, we examine the potential ambiguity of some types of references along with strategies to avoid communication failure.

Results and Conclusions: We identified 10 distinct ways in which links from text to visuals can be established. Line drawings were referred to more than visuals showing display content. The clarity of the link between text and visuals may be affected by the use of spatial expressions (such as *left/right*) that presuppose an underlying perspective, as well as by the use of semantic and functional information that is not readily accessible to laypersons. Ambiguity can be avoided by using clearly defined labels, specific perceptual information, and by repetition of visual elements in the text. Also, overspecification can be useful for enhancing communication. We conclude that technical writers need to attend to the links between text and visuals so as to support the users effectively, avoiding ambiguity by purposeful strategies. Because our study is qualitative and analytic, implications are limited by the scope of our study, and by the lack of empirical user comprehension studies.

List of index terms: household appliances, visual information, instruction guides, user manuals, consumer information

INTRODUCTION

In today's technologized world, it is common to be surrounded by technical equipment, ranging from mobile phones and computers to household appliances such as stoves and washing machines. The setup, operation, and maintenance of unfamiliar appliances is supported by manuals and operating instructions, which typically include various visuals such as pictures, diagrams, line drawings, screen captures, photographs, or pictograms. In text

types that mainly aim to entertain and engage, such as magazines and children's books, visual illustrations serve as eye catchers, raising interest and engaging the readers' imagination.

Operating instructions, in contrast, are meant to support customers in using their new appliance in a maximally efficient way. Accordingly, visuals should be systematically employed so as to serve this purpose as optimally as possible. In particular, it is essential to enable the reader to maintain a clear representation of the precise connections, not only between the textually and visually presented information but also to the real world, especially to their own instance of the appliance described in the manual. In this paper, we raise the question as to how textual and visual information is combined to guide the user's attention focus precisely towards the relevant parts of a household appliance. The challenge, we propose, lies in providing sufficient hints about the relation between textually presented information and visually illustrated configurations to enable the reader to make the transfer to the real world and apply the instructions correctly. In this process, potential ambiguity needs to be recognised so as to avoid any potential miscommunication and conceptual problems.

Previous research on multimodal texts and text comprehension (as discussed in the next section) has provided useful explanations as to why visuals are included so widely in operating instructions, highlighting their positive effects on the users' interpretation. However, beyond pointing to the spatial proximity of visuals and texts, little is known about the range of ways in which connections between both are established in existing manuals to guide the reader's attention focus. In this paper we provide an exploratory qualitative study of operating instructions for two household washing machines by the international company Miele & Cie. KG. Our analysis addresses both the German and the English versions of operating instructions to provide a basic comparison between languages. Our research was guided by the following questions:

1. How is textual and visual information combined in operating instructions so as to guide the user's attention precisely towards the relevant parts of a household appliance?

2. In what ways can ambiguity arise, and what kinds of strategies can be used to avoid miscommunication and conceptual problems?

This study presented in this paper addresses these questions. First we situate the study in the literature. Then we explain the methodology underlying our study. Afterwards, we present the results and close by offering conclusions, discussing the limitations of the study, and suggesting future research.

LITERATURE REVIEW

We discuss previous insights on the interplay of text and visuals in the area of reading comprehension in the context of user manuals, focussing in particular on the vital importance of efficient links between both representation modes. This section starts with an overview of our theoretical orientation followed by our selection of literature for the review. Then we turn to the challenges involved in operating instructions, and finally discuss visual representations in operating instructions including their various forms and the interrelations between text and visuals.

Theoretical Orientation

In this paper, we adopt a theoretical orientation that combines linguistically motivated discourse analysis with insights from cognitive science. As linguists, our primary expertise lies in understanding how communication works in relation to the linguistic and semiotic choices made by writers and speakers. One major tenet in linguistically motivated discourse analysis is that such choices are immensely meaningful; even subtle differences in linguistic expression and accompanying visualizations can have substantial consequences, and they typically express different functions or cognitive aspects [1], [2], [3], [4]. A close examination of patterns in linguistic choices frequently aims at identifying the features of a register [5] or genre [6]. In this paper we take a step further in that we do not only aim to identify the existing features of a particular aspect of a professional genre— textual references to visuals in operating manuals—but also take a critical look at the implications of these features in

terms of communicative efficiency. For the latter purpose, we mainly draw on a second relevant area of expertise, namely cognitive science, which contributes insights on text comprehension [7], [8], [9], [10], and cognitive processing during problem solving. In our context, learning to operate an appliance can be understood as a problem solving process for which a manual is supposed to provide assistance – and this may be more or less successful, depending on its features [11],[12],[13],[14].

Selection of Literature for the Review

To address our concerns in this paper, our literature review starts from previous accounts in the area of professional communication, focusing in particular on analyses of illustrated operating instructions addressing communicative and cognitive challenges. To gain deeper insights, we follow this up with literature extending beyond the realm of this genre, addressing in more detail the cognitive aspects pertaining to the comprehension of textual and visual information. This includes fundamental insights on reference resolution, which is particularly relevant considering the manual reader's task of relating the textual information to an associated visual illustration and further to the actual appliance, through resolving references in the way intended by the technical writer.

Challenges for Operating Instructions

Operating manuals are usually procedural instructions that tell the user how to set up an appliance, how to operate and maintain it, and how to solve any problems. For these purposes, they are meant to be used in relation to the actual appliance, that is, the user is expected to switch between the operating instructions and the appliance. The multi-modal and multi-informational context of operating instructions leads to a high degree of complexity with respect to cognitive processing. It involves not only the instructions presented in the operating instructions and the equipment the user wants to use, but also the user's prior knowledge about this or similar equipment, and further cognitive characteristics such as personal ability and working memory capacity [11].

If manuals are to be supportive and useful in light of this complexity, they need to correspond to the user's expectations and cognitive requirements [12]. Explanations and instructions need to be conveyed to the reader in a cognitively adequate way, satisfying their needs at any given point in time and avoiding failures of communication between the technical writer and the reader. This can be problematic because the interacting parties are involved at different times and distant locations [15], [2]. As a consequence, writers do not know any details about prior knowledge on the part of the readers (which may lead to superfluous and therefore potentially annoying information), nor do they know if the readers have problems understanding the instructions (implying that more explicit information may be necessary). In both respects, technical writers make informed guesses based on their experience, and they may adhere to guidelines (such as [16]). Readers, in turn, cannot signal that a part of the instruction has been understood, nor ask for clarification if needed [15], nor contribute to the communication by providing complementary information reflecting their collaboration in achieving a joint goal [17]. Furthermore, because both parties are in different places they cannot rely on visual cues such as gestures [18], which are central for communication about spatial configurations and relations [19].

Because none of these common devices for ensuring successful communication in natural face-to-face dialogue are available [20], other solutions need to be found if serious misinterpretation is to be avoided. This is especially crucial because of the need to refer to objects – the appliance and its parts – in the real world. The writer has good reason to assume that the reader currently has visual access to the appliance that the manual refers to. However, the spatial relationship between reader and appliance is unknown, and physical pointing gestures are patently not an option. Therefore, it is necessary to refer to this centrally relevant physical object, or specific parts of it, using other methods and representation devices. Visuals are suitable for this purpose because of their basic features: they are analogical representations of the object that they depict, and, often, they are schematic. Both of these features are

essential for human cognition in that they support focusing on those aspects of a situation that are relevant and required for reasoning, rather than distracting the reader through inessential details [21],[22]. In fact, it has been claimed that some kinds of schematic depictions may represent human conceptions of reality rather than reality itself [23]. Thus, visuals that resemble the actual appliance, or parts of it, can be very efficient features of operating instructions if employed sensibly. For instance they can help the user to recognize the elements referred to in the text on the actual appliance [24].

Language can be said to achieve similar effects in different ways, namely by schematizing otherwise complex situations [25], [26]. For example, a simple term like *over* suggests a complex spatial pattern, often (but not always) involving movement, as in *over the fence*. However, in operating instructions for household appliances, the schematic nature of language as such is insufficient to ensure successful communication because some reference items may be hard to describe in words only. Moreover, even simple actions (like pressing a button) can trigger complex mechanical procedures. If language and visuals are to be used as pointers to complex real-world objects such as washing machines, such references need to be clear, unambiguous, and specific enough to establish the relevant relations and provide the cues needed for action at any given time. Crucially, textual and visual information need to be sufficiently interrelated to each other so that any textually given information can be connected to the corresponding visual part [16], which, in turn, should simplify the mapping to the real world [24]. In the following, we take a closer look at the literature highlighting the role of visuals in operating instructions.

Visual Representations in Operating Instructions

It seems that the vast majority of operating instructions are multimodal in that they include visuals of some kind [27]. Maes and Lenting [13] observed that 95% of the operating instructions found in the authors' homes included instructive visuals. Visuals can be used in operating instructions for a number of reasons: they have a motivating effect on readers, they

offer visual relief on pages full of text and, most importantly, they are said to facilitate comprehension and improve the effectiveness of user actions. A number of studies have shown that mixed formats of text and visuals lead to a better understanding and performing than just text or visuals alone [11].

While pictorial representations of procedural instructions were once seen as an alternative to textual instructions for users with poor verbal skills and as a way of providing multilingual instructions [11], research has shown that visual representations of instructions can be just as difficult to interpret as textual instructions – especially if they come without textual elaboration. First of all, to understand visual instructions, users have to learn to "read" drawing conventions to correctly interpret the visuals [23], since drawing conventions can concern scale, temporal sequencing and movement in three dimensional space [11], and they can differ between cultures. Secondly, visuals are limited: depending on the scenario it can be very difficult, if not impossible, to clearly express goals of instructions, conditional steps, temporal relations, and differences between steps [13]. This is related to the fact that a visual in a printed manual is necessarily static, while the content depicted and described may be a dynamic process that involves movement and changes in the composition of elements [14]. To achieve optimal understanding, adapting a particular perspective on the scene may be decisive [28], which may not in all cases be straightforward to achieve. For all those reasons, pictorial representations are often regarded as unsuitable as the exclusive medium in operating instructions. Nevertheless some companies (such as the Swedish IKEA group) do favor this method, in part actually serving as a marketing strategy based on the language independency that comes with a sole reliance on visuals.

A combination of text and visuals provides opportunities for taking advantage of both by balancing the other mode's limitations. Unlike text, visuals resemble the actual appliance or parts of it and thus make recognition of objects in reality easier for the user. They are a better format for explaining location or spatial information, whereas text often gives less

precise or ambiguous information about location [13]. In addition, visuals can function as a memory aid in that seeing a picture can re-activate the representation of information that has been read and processed before [14]. Crucially, visuals can support the reader in building up a mental representation through a better understanding of what the writer had in mind when formulating the text – provided that a clear connection between text and visual is available. According to Paivio's theory of dual representation, verbal and non-verbal cognitive systems closely interact in the interpretation of material from diverse sources [29]. Glenberg and Langston [10] even suggest that the mode in which information is transmitted may not make a difference on the cognitive level, since one single mental model is built up from any resource available. The mental model itself is thought to be spatial in nature, in particular when spatial configurations are represented. However, even abstract relationships may be processed and transferred into a mental model that has quasi-spatial features [9]. At any given time, this mental model represents the reader's understanding of the information accessed so far. It is updated during the reading process by adding or deleting certain elements, corresponding to the current focus or topic of the text, and also to the state of the described object. For example, an update of the mental model occurs when noticing a new element that is relevant at a specific stage during a procedure [10]. Therefore, noticing an object that is depicted visually close to the text can be crucial for guiding the reader's attention.

Visuals can assist in constructing a mental model because the composition of elements and their interrelations resemble the required structure of the mental model, provided that text and visuals represent the same facts or situation. Visuals can repeat information from the text, enabling the reader to check whether their mental model is correct [14]. Visuals can also add information by spelling out relationships that are only implicit in the text and would have to be inferred by the reader [9] requiring a higher cognitive load (leading to longer reading times) than if visuals are included [11].

Clearly, therefore, the combination of text and visuals has a positive effect on the

readers' understanding, although the creation of a mental model as such cannot be directly observed. For further exploration of the benefit of these findings for household appliance manuals, it is crucial to gain insights about the precise ways in which such combinations highlight the relevant relations to real world objects in order to achieve specific effects. For this purpose, it is useful to look closely at the ways in which the relationship between text and visuals can be established.

In the following, we take a closer look at the diversity of forms of visuals in operating instructions, and then provide an account of possible relations between text and visuals that support these functions, as far as these are derivable from the literature. Our study of washing machine manuals then serves to explore these issues in more detail and highlights the points where referential ambiguity may hamper usability.

Forms of Visuals in Operating Instructions

Like other visuals that accompany texts, visuals in operating instructions can be essential to the understanding of the text or facilitate it. They can direct the users' focus of attention towards relevant aspects, as well as illustrate consecutive steps during operating procedures. Different forms of visuals can serve to support these functions in different ways. Visuals can provide an overview of the appliance, they can point to important elements by graphic contrast or by arrows or symbols, and they may show a close-up of certain objects or parts of the appliance [30]. In procedural instructions, arrows or crosses may be used to indicate movement [31]. Some depictions show a person carrying out an action, ideally from the point of view of the performer of the action – although this may lead to representation difficulties [28]. Photographs and realistic drawings can convey a large amount of information, but they also include visual clutter that may distract from the current purpose instead of focusing the reader's attention. Line drawings, in contrast, make it possible to remove unnecessary details, leaving only the information relevant for effective understanding while still representing real objects sufficiently to be recognizable and useful [23]. Furthermore, the content of a display

of a household appliance can be represented, for instance by showing a required start screen or the outcome of an action step. Similar to screen captures in computer manuals, this allows the user to verify the display and thus facilitates action planning, error rectification, and reassurance [32].

Interrelations Between Text and Visuals

Kress and Van Leeuwen [27] observed that elements that are perceived as disconnected will be interpreted as being independent from each other, whereas elements that are perceived as connected will be understood to be belonging together in a continuous or complementary way. But how can a connection between text and visual be established; how can a writer convey just how they relate to each other?

First of all, clearly spatial proximity is important. The reader of the document should be able to focus on both the text passage and the relevant visual without having to turn pages or having to switch between distant locations on a page. According to [33], learning improves when multiple sources of information are presented in a spatially integrated format rather than separately, because the users do not need to split attention between both sources of information or search for specific information in two different places.

Secondly, references in the text to the visuals play an important role. References to visuals can lessen the risk of miscommunication when communicating across times and locations [2], since they serve to replace pointing gestures in directing the reader's attention, either to the visual in general, or to specific elements illustrated in the visual that in turn serve the function of pointing to parts of the real-world appliance. Like the variety observed in visuals in the previous subsection, textual references can come in various forms. The connection between text and visual can be created through a repetition of visual symbols, elements or words in both text and visual, or it may be created in a less explicit way through textual cues [13], [33]. This variability will be explored more systematically in our study presented in the next section.

Thirdly, coherence between text and visuals is established on the level of content, by the distribution of information between them [30]. If information is directly repeated by illustration in a visual that has been described in the text, this may be perceived as sufficient for the reader to establish the connection simply on the basis of congruence. Similarly, information is understood as complementary and thereby potentially relevant by association if the text includes abstract descriptions that are complemented by the elements shown in the visual, or when the visual can only be understood by the complementary information given in the text. However, if text and visual are only connected through content and via inferences on the part of the reader, establishing the relevant relationships is much harder; in some cases readers will not be able to understand text or visual without valid prior knowledge [30]. Erroneous inferences made by the reader may lead to substantial misconceptions and interpretation errors [12].

Considering the issue of repeated and congruent information between text and visuals, the question arises as to whether redundancy of information may enhance or hamper interpretation and learning processes. According to [13], redundancy is essential particularly in multimodal formats for learning situations and in executing procedures because it reduces cognitive load; non-redundant instructions would require constant switching between media and complicate the interpretation of text and visuals as a coherent whole (see also [34]). A further argument is that, just like explicit verbal repetition, redundancy in text and visuals require the reader to process information twice, which may facilitate comprehension and memory [10]. Although the effect of double processing could be expected to impose additional and unnecessary cognitive load [35], empirical evidence indicates that comprehension is in fact facilitated by repetition just as well as by complementary information [10]. As observed by [36], writers of instructive texts anticipate the readers' need for redundancy, for instance by overspecifying references to visuals in the text.

While research so far has thus established convincingly that the interplay of text and

visuals is essential for a thorough and efficient comprehension of instructions in operating manuals, to our knowledge the ways in which this interplay is achieved in actual manuals on the market have not been explored systematically. A comprehensive account of the options available to establish these relations can support the identification of potential traps and dangers that instruction writers should avoid, for instance by using types of references that are clear and unambiguous, and by providing a sufficient amount of redundancy for users to support the intended inferences with respect to the real-world appliance in front of them. In the following, we present our analysis of the relation of text and visuals in washing machine operating instructions provided by the global manufacturer Miele.

METHODOLOGY

Here we provide the methodological background required for understanding the procedures and outcomes of our study. Following an outline of our choice of a research methodology, we explain how we sought our participants, how data was collected, how data was analyzed, and report on how we assured credibility and trustworthiness. We address the following research questions:

1. How are textual and visual information combined in operating instructions so as to guide the user's attention focus precisely toward the relevant parts of a household appliance?
2. In what ways can ambiguity arise, and what kinds of strategies can be used to avoid miscommunication and conceptual problems?

Choice of a Research Methodology

Most research addressing textually and visually presented information focuses on empirical studies carried out in a laboratory to assess efficiency and comprehension using controlled settings. We, however, were interested in the features of methods actually used in current appliances, along with the manufacturer's associated policies, and an informed assessment of the clarity of these features. Therefore, to address our first research question we obtained relevant background information concerning the design process from our participating

manufacturer, and we explored in detail the features of text-to-visual reference types adopted in a sample of operating instructions that represents a common type of manuals. To address our second research question we re-examined literature relevant to the various identified reference types so as to identify possible sources of ambiguity and miscommunication. This provided a basis for identifying strategies to avoid misunderstandings in this area of professional communication.

Participants

We sought a company that would provide access to its technical writers and would allow us to evaluate their consumer documentation for appliances. The company should be well established in the area of everyday products, ensuring that the operating instructions were accessible to lay users, professionally designed, and based on extensive experience across various products rather than produced on an ad-hoc basis. Also, international recognition and the existence of multi-lingual manuals were desirable, enhancing the global significance of our study. We approached our chosen company in person, without offering any incentives for the study. Ethical approval was not required for this type of study, according to the standards of research practice in Germany.

How Data Was Collected

Here we describe our procedures of collecting data that enabled us to address our research questions in an in-depth qualitative and systematic way. This section starts with Interviews, followed by Document Analysis.

Interviews

Since our first research question aims at identifying strategies used in operating instructions for combining texts and visuals, we were interested in obtaining information about any explicit principles and strategies applied for the design of operating instructions within the company, and in the extent to which writers independently agreed on these strategies. Accordingly we conducted short written interviews with two groups of technical writers located at different production sites. The six interview questions concerned the

illustration of operating instructions and the relation of text and visuals in the operating instructions. See Figure 1 for the precise interview questions.

INTERVIEW QUESTIONS

1. Gibt es Leitlinien bei Miele wann und wie viele Bilder eingefügt werden sollen?

Are there guidelines at Miele when and how many visuals should be inserted?

2. Wenn nicht, nach welchen Kriterien wird entschieden, wo und wie viele Abbildungen eingefügt werden?

If not, which criteria are decisive for how many visuals will be inserted, and where?

3. Wie wird an die Konzipierung einer Abbildung herangegangen? Nach welchen Aspekten wird über die Gestaltung einer Abbildung entschieden?

How is the conception of a visual approached? Which aspects determine the composition of a visual?

4. In welchen Fällen wird z.B. eine Person oder eine Hand in eine Abbildung eingefügt?

In what cases is, for instance, a person or a hand inserted in a visual?

5. Wie wird das Bild für gewöhnlich in den Text eingebunden?

How is a visual usually embedded in the text?

6. Seht ihr die Abbildung eher als unterstützendes Element, das einen Punkt verdeutlicht, der bereits im Text erklärt wird oder ersetzt es Informationen, die sonst im Text erläutert werden müssten?

Do you see visuals more as a supporting element that clarifies a point already stated in the text, or do visuals replace information that otherwise would be needed to be included in the text?

Figure 1. Questions used in the interviews with technical writers

Document Analysis

Since we aimed at an in-depth understanding of the features of links between text and visuals, we restricted our analysis to a small but representative set of operating instructions for appliances by a global company, randomly selected out of a recent publicly available catalogue. In order to focus on the features of these instructions rather than complicating our analysis by content-related diversity, we allowed for variability only with respect to two dimensions relevant for our research: complexity of the appliance, and choice of language. This yields a total of four related operating instructions that are representative of their kind, while differing in ways that are common for household appliance manuals.

How Data Was Analyzed

By examining the content of the collected interview responses we aimed to gain relevant insights that helped narrow down our selection of operating manuals, and also provided initial answers relevant to our first research question. Furthermore we attended to consistency between the participating groups of technical writers.

To address our first research question systematically, we proceeded in two main steps, combining in-depth qualitative analysis with a look at quantitative distribution patterns. In particular, we aimed for a systematic account of the types of textual references to visuals in our chosen cases of operating instructions with respect to a) their variability (in terms of the repertory of communicative devices used to establish reference), b) their distribution in our four sets of instructions, and c) their interplay with different types of visuals. In the process of identifying and classifying types of references in the operating instructions, we attended to the inclusion of graphic elements, specific words or word groups, repetition of words or visual elements, and specific sentence structures. Classification was done first by one of the authors and then checked by the other author. Following classification, we identified their quantitative distribution across our data, relative to the dimensions of complexity and language, and relative to the type of visual to which a textual reference belonged.

To address the issue of generalizability, we informally inspected other operating

instructions for various types of appliances in order to verify the existence of similar reference types across instances of this genre, as well as to gain an impression of their distribution.

While this informal part of our study is too limited to warrant any conclusive generalizations, we aimed to confirm that our exploratory qualitative approach indeed captures common rather than exceptional phenomena and is thus of relevance to a wide range of professional communicators concerned with operating instructions.

Based on this systematic account of observed variants along with their distribution, our second aim in this paper is an evaluation of the relationship between text and visuals with respect to potential ambiguity, using the principles and patterns identified in our data as an example. In this respect, our second research question was addressed by examining possible problems and strategies to avoid miscommunication and conceptual problems, based on relevant insights drawn from the literature.

Assuring Credibility and Trustworthiness

Our contribution triangulates various perspectives on the issue at hand, by combining reports from technical writers with an in-depth analysis of the features found in operating instructions, and further discussing them in light of a wealth of literature that directly relates to the identified features. Readers who identify similar design principles in other manuals can therefore use our analysis as an example in order to assess the significance of these design principles for usability.

RESULTS

Here we present the results of our qualitative study addressing types of links between pictures and text in our data set, and demonstrate how these types are distributed. This serves as a basis for evaluating the various reference types with respect to potential ambiguity and proposing effective linking strategies. This section starts with an account of who participated in the study, followed by results of the interview; then we address both of our research questions in turn.

Who Participated in the Study

Based on our desired selection criteria, we were able to win the company Miele for participation in our study. Miele is a German manufacturer of high-end household appliances including washing machines, tumble dryers, and kitchen appliances such as stoves, ovens, steam cookers, coffee makers, coolers, freezers, and dishwashers. The company is based in Gütersloh (Germany) but is in business internationally; the company's products are sold in roughly 100 countries worldwide.

The interviews were conducted with technical writers working for the Miele company, who discussed the answers to the written interview questions as a team before sending them back to us. One of the groups (consisting of two experienced technical writers) was located at the Miele production site in Gütersloh, while the other group (consisting of three experienced technical writers) was located at the production site in Bielefeld (Germany).

We identified the operating instructions in English and German for two different washing machines for the purpose of analysing reference links. One of the washing machines (model W1914) was a basic model with a limited range of actions, and the other (model W5964) was a more elaborate version with more features and settings. The German and English appliances, and accordingly the basic content of both sets of operating instructions, were identical except for one detail. The German appliance had a so-called water-proof system that was not part of its English appliance counterpart. As a consequence of this technical difference, the German version of the operating instructions includes one more visual than the English version.

Results of the Interviews

We could gain the following insights from the technical writers' responses.

As both groups confirmed, there are no explicit guidelines for technical writers at Miele that prescribe how visuals should be included. Nevertheless, the answers from both groups were generally remarkably similar, although they answered the questions independently, and produced operating instructions for different kinds of appliances. Some unwritten rules and

patterns emerged. For Miele technical writers, visuals have a leading or text supporting function; additionally one group stated (quotes translated from the German original): "There are however documents, e.g. assembly instructions, where the visuals replace the text". The decision about when and where to add a visual depends on the complexity of the action, as one group elaborates: "How complex an action is, and/or whether several steps need to be carried out. A visual always belongs to an action instruction, and is always placed above it. No more than two-three steps are explained in one visual. In case of more steps there will be further visuals." The other group adds: "[depending on] the concept of the document (mono- or multilingual, text-centric (as in operating instructions) or graphics-centric (as in assembly plans)." The instructions analyzed here all belong to the monolingual text-centric operating instruction type. This supports our decision to use a small, homogeneous and manageable data set that is representative of its kind for an in-depth investigation of its features and their implications.

The interviews further revealed that, after deciding to add a visual to a particular instruction, a photograph is taken of the relevant part of the appliance. The photograph needs to provide a suitable amount of visual context to enable the user to identify and locate the relevant part, and it needs to employ a suitable perspective on the scene. Then a graphic designer renders a schematic drawing of the photograph, omitting unnecessary details and highlighting the most relevant parts of the scene. Symbols are added such as arrows (formerly: hands) showing movement, a magnifying glass to emphasize a detail, warning symbols, and further information such as measurements in the form of text. It is important in this process to ensure consistency within a document: "adjustment to the concept of the other visuals in the document (same way of depicting arrows, shading, ...)". The visual is then included at a fixed position, which can differ across appliances; for washing machines, the conventional position is above the text describing the relevant action step.

Due to the company's German base, operating instructions are usually written in

German and then translated into other languages as needed. Translation processes such as these always involve minor changes in content, so as to adapt the operating instruction to the local culture and the local technical regulations and safety regulations. The visuals remain the same, unless they include language-specific content; in that case language-specific elements such as captions are adapted to the target language.

These insights provide the conceptual background needed for our study. While there is no explicit set of instructions for links between text and visuals, technical writers nevertheless adhere to various well-established procedures that lead to the patterns and principles we address in our analysis. The fact that translations are not necessarily verbatim (though remain closely related in content) while visuals are identical allows us to investigate potential subtle differences in the different language versions. We were interested, on the one hand, on the German original version (since Miele is based in Germany), but also wished to examine a data set of relevance for a wider audience, closer to other English-language based studies (such as [7], [10], [13]). The principles adopted at Miele allowed us to do a straightforward parallel analysis of similar versions in these languages.

Answers to Question 1: How are textual and visual information combined in operating instructions so as to guide the user's attention precisely toward the relevant parts of a household appliance?

In the following, we will define a qualitative categorization of prominent types and features of textual references that we found in the operating instructions. This categorization serves as a basis for a quantitative analysis of the distribution of reference types with respect to the following distinctions:

1. Basic vs. advanced washing machines
2. German vs. English versions
3. Line drawings vs. display content

Each type of reference in our categorization will be illustrated by an example taken from the English version of the operating instructions for the advanced washing machine (W5964). To

point to the part of the visual that we wish to emphasize, we highlight it by shading in our figures. Wherever possible, we discuss features of the identified type of reference in light of the available literature.

Labeled Elements in Visuals

Elements in visuals can be labeled by using numbers or letters, allowing the writer to refer back to those elements in text (Figure 2). These labels support the identification of elements in the visual, which in turn leads to the identification of the relevant parts of the actual appliance. Labels can occur in various forms. First, the label may be positioned within the visual directly next to the relevant element. Second, label and element may be connected by a reference line. Third, labels may be listed closely to (but outside of) the visual, while the visual itself includes letters or numbers close to the relevant element. This allows the reader or user to relate labels and elements with the same letter or number. Combinations of those different presentation styles are also possible [14], [33].

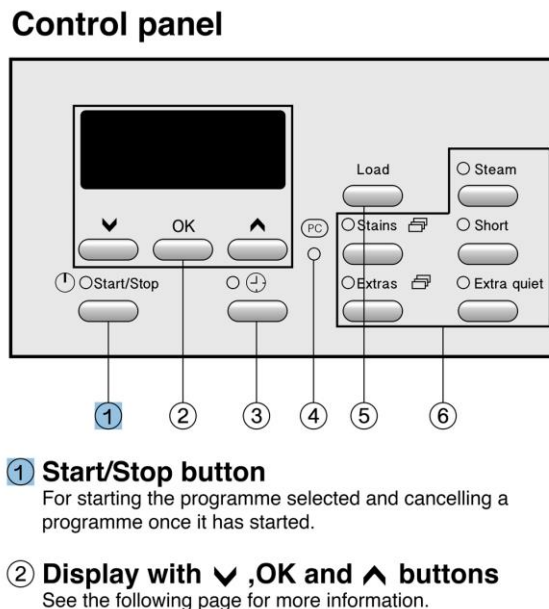
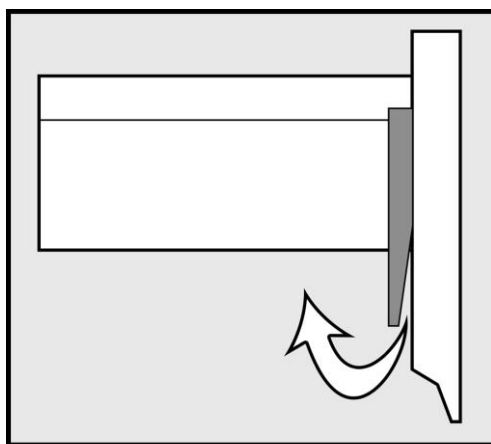


Figure 2. Reference using labeled elements

1. Dimensional Expressions The text may employ spatial expressions to describe the position of an element in the picture (and on the appliance itself) (Figure 3). Relevant expressions include *links*, *rechts*, *oben*, *unten*, *vorne/vor*, *hinten/hinter* in German and *left*,

right, top, bottom, front/in front, back/behind in English. This class of terms has been variously referred to as "projective" or "dimensional" in the literature [37]. Since these are similar but not identical in scope and semantics in the two languages, it is particularly interesting to look at the solutions in both versions in our data set.

Dimensional expressions such as these are potentially ambiguous in that they depend on the perspective taken on the scene. For instance, what a user identifies as the *left side* of an appliance may depend on the current orientation of the appliance, or of the user relative to it. Other spatial expressions like *top* or *bottom* are less ambiguous in the given context of immobile appliances, since they are defined by gravity [13]. Proximity terms like *near* or *close* are not perspective dependent but underspecified with respect to directions; however these were not found in the operating instructions at all.

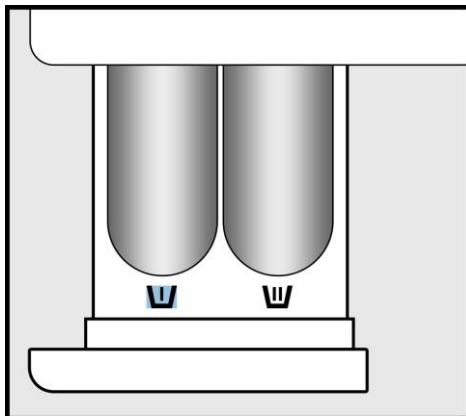


- There is a lever **behind** the detergent drawer fascia panel

Figure 3. Reference by spatial expressions

2. Repetition of Visual Elements or Symbols The visual may include a visual element or symbol that is repeated as part of the text (Figure 4). Since they are visually distinct from the text's ordinary characters (letters and numbers), they are easily detected in the text. This type of reference can create clear and unambiguous references between text and visual since there is a direct visual correspondence rather than mediation by numbers, symbols, or spatial descriptions. However, a simple repetition of a visual element does not provide any semantic

or functional information that may contribute to the understanding of the function or name of this element [14].



- Pull out the detergent drawer and add detergent to the compartments as follows:
- W Detergent for the pre-wash 1/3rd of the total recommended amount.

Figure 4. Reference by repetition of visual elements

3. Perceptual Information The text may provide perceptual information about elements in the picture, that is, refer to visible features of elements in the visual (Figure 5). This may concern the shape or color of objects or elements; for example in the case of display content the highlighted, coloured text [13].

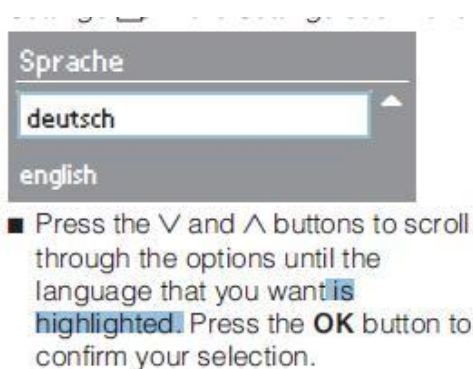


Figure 5. Reference by perceptual information

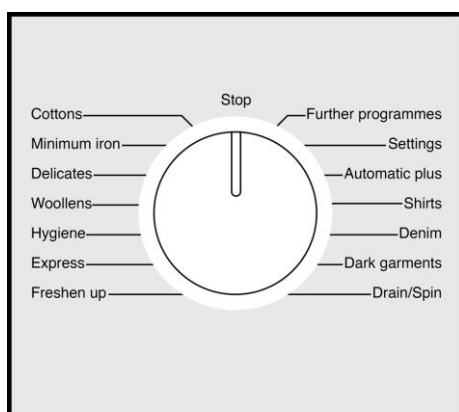
Repetition of Words Words may be displayed in both the visual and the text (Figure 6). In addition to being a rather clear and unambiguous type of reference, this kind of reference also

provides semantic and/or functional information about an element or step relevant in the given context [14].



Figure 6. Reference by repetition of words

4. Semantic and Functional Information The text may provide functional or semantic information about elements in the picture (Figure 7). When an element is named, its name very often gives information about its function. The name and/or function of an element may be available to the user on the basis of their prior knowledge, or the user may be able to draw conclusions about the function of an element based on the combination of text, visual and the appliance in reality [13].



- Turn the *programme selector* to the required programme.

Figure 7. Reference by semantic and/or functional information

5. Action Performance Certain actions are described in the text using action or motion verbs (such as *drücken*, *ziehen*, *drehen* in German or *push*, *pull*, *turn* in English). The same action can be visualized in the picture by using graphic elements (Figure 8), such as arrows [30].

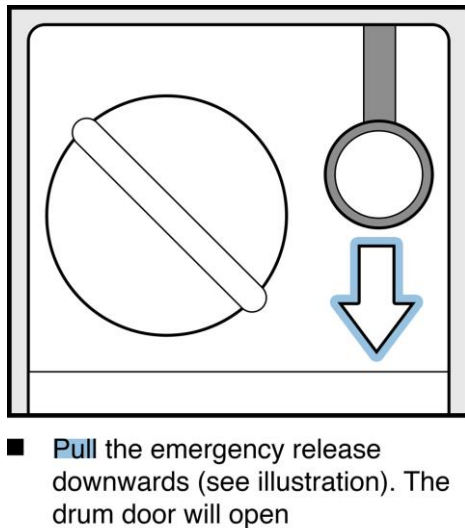


Figure 8. Reference to action performance

6. Embedding of Visuals in Text The visuals can be embedded within the explanatory text, rather than appearing as separate elements outside the text. This is done by use of a distinct sentence structure and punctuation marks, such as colons and ellipsis. The visual thus becomes part of the sentence (Figure 9). In contrast to most visuals that are not strictly necessary for comprehension, these visuals are a vital part of the instruction; the text standing alone would be useless for the user. Only in combination with the visual does the text provide useful information.

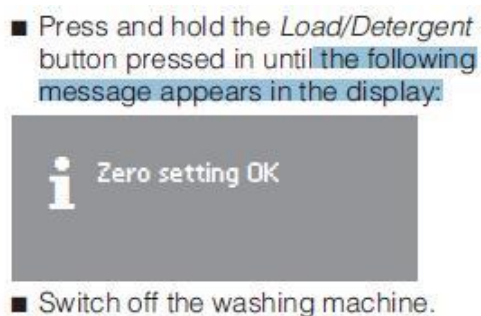
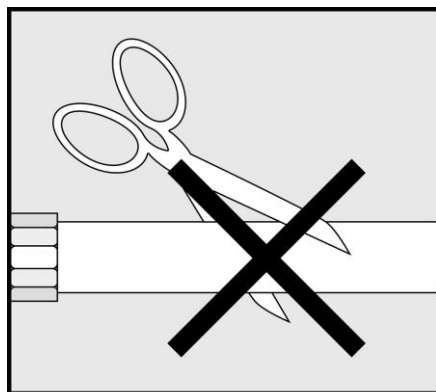


Figure 9. Reference by embedding visuals

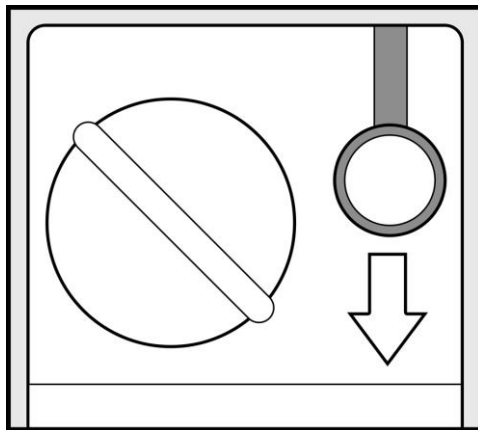
7. Prohibition The text may aim to prohibit certain actions by the use of linguistic negation forms, such as *nicht*, *kein* in German and *do not*, *not*, *no* in English. Such linguistic items naturally combine with a visual where certain elements are crossed out to symbolize the action in question (Figure 10). Visuals that convey prohibitions can be pictograms or at least use a similar style of representation [23], [30].



- The protective sleeve of the hose must not be damaged or kinked

Figure 10. Reference to prohibition of an action

8. Direct Textual Reference or Description The text may direct the user's attention toward the visual by either asking the reader directly to look at the visual (as in *see illustration*, cf. Figure 11), or by describing what the visual shows (as in *The load size indicator display will appear*). In the latter case, the effect of spatial proximity is enhanced by semantic correspondence between text and visual, but there is no direct reference.



- Pull the emergency release downwards (see illustration). The drum door will open

Figure 11. Reference by direct textual reference or description

Having outlined the repertory of types of textual references to visuals that could be identified, we now turn to their distribution in Miele's operating instructions.

9. General Distribution The overall number of textual references to visuals in the instructions was higher than the number of visuals as such, since a visual could be referred to in the text more than once. As Table 1 shows, the ratio of references to visuals was similar for all four versions of the instructions. In each operating instruction, precisely 9 of the visuals (all of which were visuals showing display content) were not referred to at all; Table 1 gives the percentage relative to the number of visuals in each instruction. On the other hand, in some cases two references referred to exactly the same element in the same visual, leading to overspecification (see discussion below). In most of these cases, both for the basic and the advanced washing machine, elements in the visuals were referenced by the reference type **labeled element in visual** as well as by **semantic/functional information**. In the operating instructions for the advanced washing machine, a combination of the reference types **semantic/functional information** and **repetition of visual elements or symbols** was also found.

Table 1. References to visuals in Miele's operating instructions

	German: Basic	English: Basic	German: Advanced	English: Advanced
Number of visuals	53	52	66	66
Number of references	73	75	94	96
Ratio references: visuals	1.38	1.44	1.42	1.45
% visuals without a reference	16.98	17.31	13.64	13.64
% visuals with over- specified references	7.55	7.69	15.15	15.15

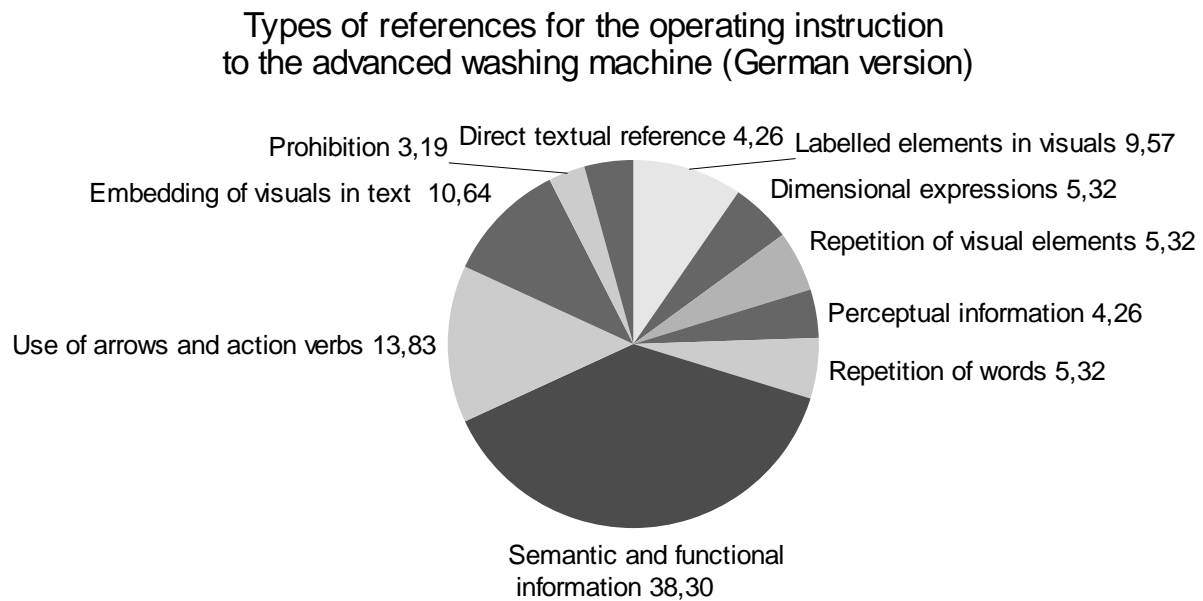


Figure 12. Distribution of references in the German advanced washing machine manual

Figure 12 shows the distribution of reference types in the German instruction for the advanced washing machine. Most of the categories introduced in the previous section are represented with fairly low frequency (below 15%). Only one category falls out of this pattern, namely the references providing semantic and/or functional information (38.3%). References to action performance (13.83%), embedding of visuals in text (10.64%), and labeled elements in visuals (9.57%) are somewhat more frequent than the remaining categories, which are represented at or below a 5.32% threshold.

To gain further insight about the distribution of reference types, we examined three factors that might influence how visuals are embedded in text: the type of washing machine described (basic vs. advanced), the language used (German original vs. English translation), and the type of visual (line drawing vs. display content).

German vs. English Versions Since the English version is a translation of the German, we did not expect any remarkable quantitative differences. Surprisingly the English versions did have two more references in each case, but there was no statistical difference (German:

mean=1.40 references per visual, SD=0.88; English: mean=1.45; SD=0.95; two-tailed t-test: $p=.70$). Neither the number of visuals not referred to at all, nor the number of overspecification differed between the languages (see Table 1). The types of references also did not differ (χ^2 : $p=.98$; $DF=9$).

Basic vs. Advanced Washing Machines The number of visuals and, consequently, the number of references to visuals was higher for advanced than for basic washing machines (see Table 1). The number of references did not differ (basic mean: 1.41 references per visual, SD=0.93; advanced: 1.43 references per visual, SD=0.90; two-tailed t-test: $p=.80$). However, overspecification occurred more often in the advanced versions (10 times each) than in the basic ones (4 times each) (see Table 1 for percentages). Although this difference does not reach statistical significance (χ^2 : $p=.074$; $DF=1$) due to the low number of instances investigated here, the higher number of overspecifications may arguably reflect the technical writer's perception of the advanced machines as more complex. This opens up a range of interesting issues to be discussed below. This seems to be supported by the fact that the percentage of visuals without a reference was lower in advanced as compared to basic machine instructions (see Table 1).

Similarly, the types of references differ between the machines (χ^2 : $p<.005$; $DF=9$), with percentage differences seen in almost every reference category (Figure 13). The overall distribution remains more or less the same in that semantic and functional information is predominant, with arrows and action verbs as the next highest category. However, as complexity increases (with advanced machines), the range of variety appears to become wider, with two categories added (perceptual information and repetition of words). Obviously, for a reliable quantification of specific differences between reference types according to type of dishwasher a much larger corpus of instructions would be needed.

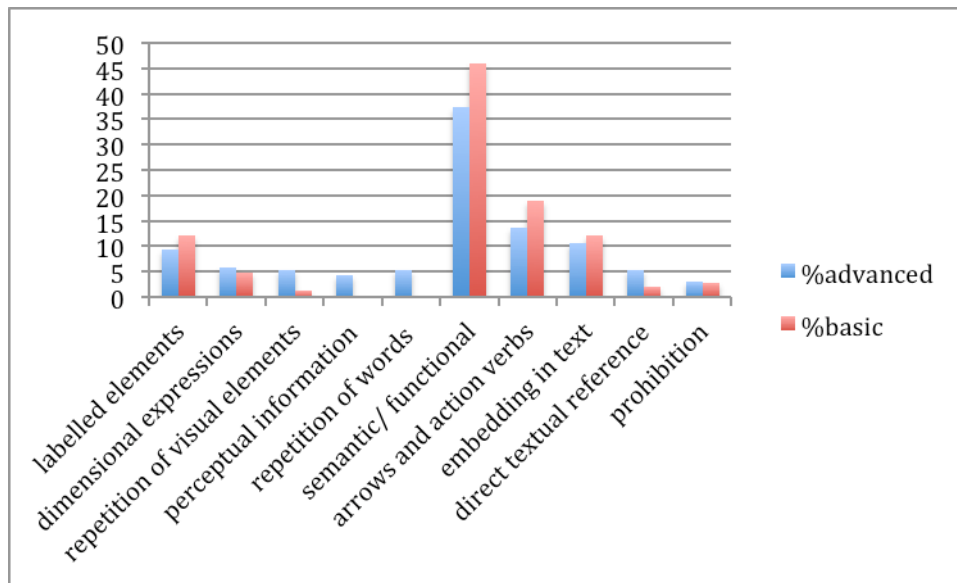


Figure 13. Distribution of reference types in basic vs. advanced manuals

10. Line Drawings vs. Display Content While our discussion so far focused on the reference types to visuals, the visuals themselves vary in a number of respects. Here we take a closer look at the basic distinction between line drawings and display content (Figure 14). Line drawings serve a number of different purposes, such as showing details and visualizing actions. Visuals that show display content typically highlight either a start screen or a goal screen in the operation of the machine. We were interested in the extent to which the type of visual affected the way in which the visuals were referred to. Since the English version did not differ significantly from the German translation we focus here on the German versions.

In both cases there were more references to line drawings than to visuals showing display content. For the basic machine, the German manual contained 36 line drawings and 17 instances of display content, and for the advanced machine, there were 37 line drawings and 29 instances of display content. Again, this difference can be traced back to the higher complexity of the advanced machine.

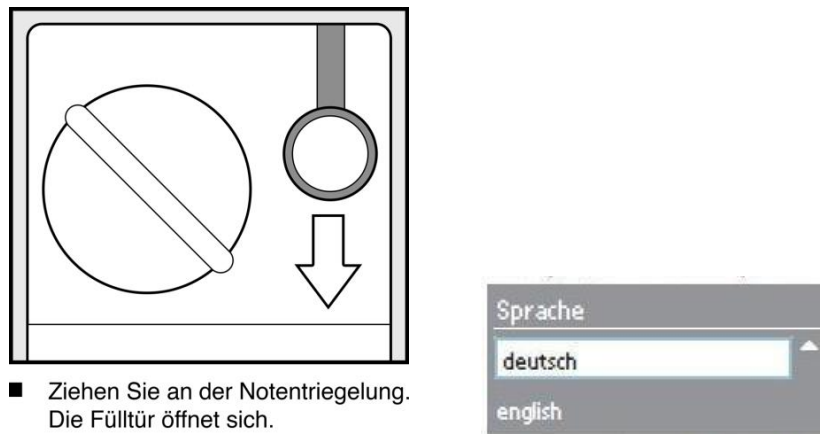


Figure 14, left: Example for a line drawing; right: visual showing display content

For the basic machine, line drawings were referred to 1.81 times on average ($SD=0.67$), while visuals showing display content were only referred to 0.47 times ($SD=0.51$). This difference is highly significant (two-tailed t-test: $p<.001$). This pattern was repeated in the advanced machine (line drawings: mean=1.89, $SD=0.70$; display content: mean=0.83, $SD=0.71$; t-test: $p<.001$). Although the difference in reference frequency between types of visuals appears to be mediated by machine complexity (since the mean for display content is numerically almost twice as high for the advanced machine than for the basic one), the difference between machine types does not reach significance (two-tailed t-test: $p=.07$).

Generalizability

As the operating instructions analyzed in this paper relate to a restricted set of appliances from one single company, we were interested in the extent to which other manuals exhibit similar features. For this purpose, all operating instructions found in the household of one of the authors (> 20 instructions including microwave oven, flat iron, radio, digital camera, notebook, and the like) were inspected informally. We found that almost all operating instructions featured at least one **line drawing with labeled elements**, typically to provide an overview of the appliance and general orientation support. Only one operating instruction used a photograph for this purpose. Operating instructions for equipment that came with some kind of display all featured visuals of the type **display content**. Depending on the complexity

of the equipment in question, especially regarding the setting manipulations, operating instructions featured a varying number of line drawings. While only one visual on the first page of the operating instruction will be sufficient for the operation of, say, a low end flat iron or radio, more complex manipulations require instructions with a stronger visual support.

As for types of references, our informal check of operating instructions for other household appliances and lifestyle equipment indicated that three of the most common types of references found in our analysis indeed seem to be typical basic types of references in operating instructions. The reference types **semantic and functional information** (which was the most prominent type in the Miele instructions) as well as **labeled elements in visuals** were found in all of the operating instructions checked. References to **action performance** were particularly frequent in operating instructions where some assembling was necessary. Other types of references found in the Miele instruction, namely **repetition of visual elements or symbols, repetition of words, and embedding of visuals in text**, were also observed, but in considerably lower numbers. Notably, just as in the Miele instructions, the reference type **embedding of visuals in text** only occurred in operating instructions that featured visuals of the type **display content**.

Answers to question 2: In what ways can ambiguity arise, and what kinds of strategies can be used to avoid miscommunication and conceptual problems?

In this section, we reconsider critically the information we have gathered and analyzed, so far, on a qualitative and quantitative basis. We now extend this analysis by evaluating the identified principles and patterns with respect to ambiguity and the avoidance of miscommunication. To accomplish this, we recur to the relevant literature on these aspects wherever feasible.

When asked about the most obvious connection between instruction text and visuals (see Figure 1 earlier in this article for the interview questions), the technical writers of Miele

uniformly pointed to spatial proximity. Indeed, in most of the cases analyzed in our study, the text right above or right below describes the situation illustrated in the visual. However, proximity alone is clearly not sufficient to represent the communicative intention of a visual. Our analysis revealed that the overwhelming majority (83%) of the visuals displayed in the operating instructions were explicitly referred to in the accompanying text, many of them more than once, resulting in overspecification (to be discussed in more detail below). References to visuals in text serve the important function of filling in potential gaps in the user's understanding by creating a meaningful, non-empty link between the visual and the accompanying text, clearly going beyond the spatial proximity aspect pointed out by the technical writers.

Research on multimodality [27] and on the processing of text and visuals in combined formats [33][11][10] suggests that the representation of information in a combined format of text and visuals can improve the processing and understanding of operating instructions. Likewise, usability testing highlights the importance of combining texts and visuals in an effective way [24]. Information provided in the text guides the reader's attention toward the visuals and specific objects in the visuals. According to the mental model theory and the split-attention effect theory, the combination of text and visuals lowers the cognitive load in processing information and allows the users to create an elaborate mental model [10] [33]. As a result, matching information from the operating instructions with the actual appliance is achieved faster than would be possible with a text-only operating instruction [11]. By enhancing the connection between text and visual, references to visuals reinforce the positive effects of their combination. As Kress and Van Leeuwen argue [27], connected elements are perceived as continuous and complementary. As a result, readers will pay more attention to visuals if those are referred to in the text rather than standing alone.

The instruction manuals we analyzed establish this crucial link between text and visual

in various ways. Across the four types of instructions analyzed, the most frequent type of reference was enhancement of the visual representation by semantic or functional information, followed by references to action performance through action verbs in the text and arrows indicating the action in the visual. Further frequent reference types were references through labeled elements and through the embedding of visuals in text. These patterns were overwhelmingly independent of the language of the manuals (English or German), as could be expected given that the manuals were direct translations appearing in closely related Western cultures. It seems plausible, however, that if a version of the instructions were written for a very different culture (such as Asian) and if sensitive adaptations were made for that culture, the use of visuals might be quite different.

However, the type of visual appeared to affect the necessity of providing an explicit link, as there were significantly more references to line drawings than to visuals showing display content. A possible reason for this could lie in the diversity of content allowed by a type of visual, related to the extent to which it appears to be self-explanatory. Line drawings can show a number of situations in a number of different perspectives, which can be decisive for the success of the communication [28]. In contrast, a visual showing display content is more limited in its form of representation. Even though a visual showing display content or screen capture can serve a number of functions in an instruction [32], it will still only show the display itself. A possible explanation of the function of this display content may then be found in the textual context. Furthermore, an instance of display content is less likely to be used for multiple processes, as each process would have different display content—that is a different (unique) system message. A line drawing, in contrast, would typically represent a component of a machine, which might be used for more than one user task. Line drawings generally allow for greater diversity of content, including close-ups of relevant objects [14] and action performing or prohibition [31]; this offers different possibilities for usage, and

opens up a wider range of interpretation. To investigate the relationship between display type and the frequency and efficiency of reference links in more depth, an in-depth study targeting usability in this regard would be desirable.

So far, we have addressed the nature of visuals and references to them, and explored the wide scope of variety found even in a small set of manuals. However, the efficiency of information conveyance may be affected not only by the existence of visuals, textual information, and links between them, but crucially by the *clarity* of this link. In the following, we take a critical look at potential ambiguity, asking to what extent the textual references allow the user to correctly and unambiguously identify the relevant element in the visual, and what strategies are employed to avoid potential misinterpretations.

Underspecification, Ambiguity, and Clarification Strategies

As mentioned above, a small but consistent proportion of the visuals in the operating instructions (see Table 1) were not clearly referred to textually; here, the only link between the visual and the text consists of spatial proximity. Notably, all of the visuals without clear references were visuals showing display content, such as a target screen, a required start screen, or an action step. In these cases, the absence of a textual link leaves it to the user to interpret at which point the display content shown in the visual is meant to occur. In Figure 15, the purpose of the visual remains unclear; there is no clear connection between text and visual. Further research is needed to explore user reactions to visuals that are not referred to in text compared to visuals that are clearly referred to in text. Some indications for the benefits of a good integration are provided by user studies such as [24].



Figure 15: Example for a visual that is not explicitly referred to in the text.

Apart from text-visual combinations that lack a concrete link altogether and are thus underspecified, other instances may involve ambiguity and lead to possible misinterpretation in more intricate ways. For instance, dimensional expressions such as *left*, *right*, *in front of*, and *behind* are widely recognized (and have been extensively studied) as being highly context dependent, with a high potential for ambiguity. Maes and Lenting [13] discuss the use of dimensional expressions in instructions, based on a categorization of three possible frames of reference, which they call deictic (viewer-centered), intrinsic (object-centered), and extrinsic (environment-centered) (see also [38] for a similar and currently more established categorization). Extrinsic frames of reference are typically not ambiguous as they are "determined either by global discourse conditions, such as gravity [...], cardinal wind directions [...] or by local salient features in the environment of the communication" [13] (p.102). In contrast, deictic and intrinsic frames of reference are potentially ambiguous, as they presuppose a particular perspective. For instance, a reference to the "front of the device" might refer to the user's current view on the scene, or to the intrinsic front side of it. This can lead to confusion if the perspective used in an instruction is not made explicit and the deictic and intrinsic perspectives do not coincide. By "front of the device", the writer may be referring to the part of the device that is closest to the user in that particular moment, or to the part that constitutes the front of the device. Adding to the complications, the intrinsic "front" of the device may not be unambiguous as such, as it is necessarily based on the user's concept of this device. It can be expected, in the case of a washing machine, that this is the side that includes control panel and drum door.

Within the operating instructions by *Miele*, the dimensional terms *left* and *right* are rarely used, but the terms *front* and *back* occur with references to the intrinsic front and back of the appliance, which is not necessarily the side of the washing machine that the user stands

closest to or farthest away from. To clarify this reference, the Miele operating instructions use visuals showing elements that are distinct to the relevant side of the appliance (such as a front door handle). The dimensional expressions *top*, *bottom*, *up*, and *down* are used more frequently. Even though these expressions are unambiguous because they are defined by gravity, they are still typically clarified by illustrations (see Figure 16, left). In some cases, directionality is only expressed visually, potentially reducing clarity in this respect (see Figure 16, right). Furthermore, note that only the English version has the additional pointer directing attention to the visual ('see illustration').

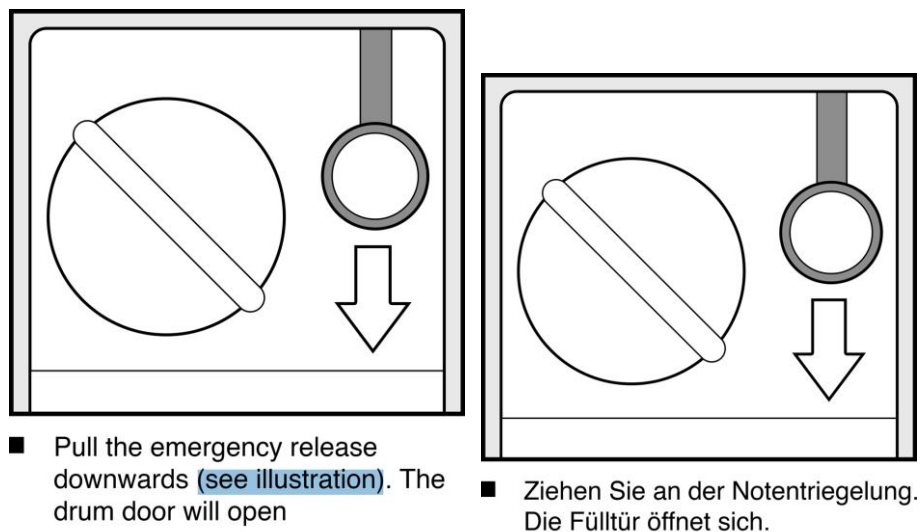
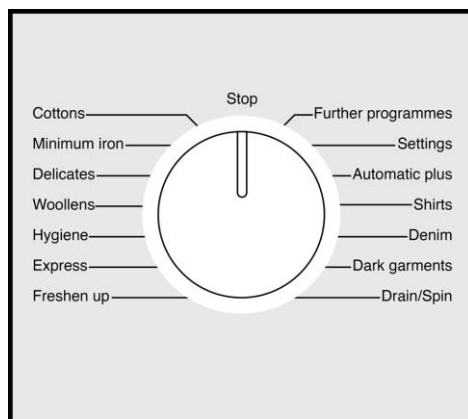


Figure 16: The directionality of the dimensional term 'downwards' in the English version (left) is clarified by the arrow in the line drawing, establishing a direct link between the text and the visual. This direct link is missing in the German version (right), where the text reads: "Pull the emergency release. The drum door will open."

Maes and Lenting [13] suggest that references to object characteristics, such as semantic or functional information, can help avoid the use of potentially ambiguous dimensional expressions. Indeed, as outlined above this is the reference type found most frequently in the Miele instructions – arguably an ambiguity avoidance strategy. However, the use of semantic and functional information is only unambiguous if the elements referred to

are known to the user. For instance, to "turn the programme selector", the user needs to know which knob serves as programme selector. Relevant object characteristics can be available to the user on the basis of specific prior knowledge (such as knowing what programme selectors on washing machines typically look like), or general world knowledge (knowing that to turn a selector, there must be some knob suitable for turning). If the user neither knows the name of the relevant element nor its function, correctly identifying the element through semantic and functional information can become a challenge. Especially in technical contexts, the correct name and/or function of elements will not necessarily be known to a layperson. To resolve this problem, labeled visuals help assign the semantic terms to the relevant elements, thus allowing a correct identification of the entity referred to. This can be done by explicitly labelling all functional elements, or by referring to them in a text-visual pairing such as the one shown in Figure 17.



- Turn the programme selector to the required programme.

Figure 17: The technical term used in the text is implicitly clarified by the visual, allowing the user to infer that the knob shown in the visual is the element referred to as "programme selector".

Perceptual elements that direct the reader's attention to relevant aspects in a visual are typically inherently unambiguous, such as references through labeled elements to action

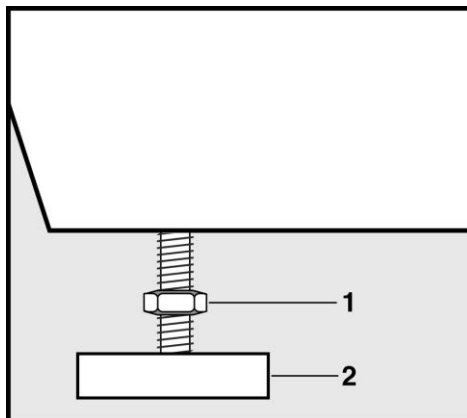
performing and prohibition [39]. References using the repetition of visual elements or symbols or words in both the text and visuals also create a clear connection between the text and the relevant part of the visual. Even though references through the embedding of visuals in the sentence and direct textual references, such as "see illustration" (Figure 15), do not direct the reader's attention to specific parts of the visuals that are most relevant at a given moment, they still are highly effective in guiding the reader towards consciously perceiving the visual.

One further well-established strategy to avoid ambiguity is overspecification: providing more information than strictly necessary for the correct identification of an object or element. Since technical writers and the user of an operating instruction are not present in the same location, the user cannot ask for clarification if unable to identify a reference given in the operating instruction. At the same time, the technical writer cannot expect direct feedback. Against this background, technical writers may tend to include more information than strictly necessary to make sure that users are able to interpret their referential expressions correctly; this results in overspecification.

Generally, corresponding to our analysis above and following [13], distinguishing information used in an referential expression may be object information, such as perceptual (shape or colour) or location information (where is the object or element located?), as well as functional information (what does this object or element do? What is it for?). According to [15] these types of information are often combined in instructive texts, leading to frequent overspecification that can amount to more than half of the references in a manual, especially for critical subtasks. According to their findings, the redundancy of an overspecified expression can speed up the identification processes in communication. Redundancy contributes to making signals more robust; readers may need more than the minimum of information to quickly and correctly identify an object or element.

Although the Miele operating instructions analyzed here appear to be somewhat less

rich in this respect, this ambiguity avoidance (and cognitive support) strategy is nevertheless in use. As reported above, about 7.6% of the visuals in the instructions for the basic washing machine and roughly 15% of the visuals in the instructions for the advanced washing machine were referred to using overspecified references. The higher number of overspecifications in the operating instructions for the advanced washing machine can be explained by its higher complexity (more setting options, and an added display). In our analysis, overspecification meant that two references referred to exactly the same element in the visual, even though one reference should provide enough information to allow the user to correctly identify the relevant element. In the case of the basic washing machine, all overspecified references referred to line drawings and were a combination of references through labeled elements in visuals and references providing semantic and/or functional information about the element (see Figure 18).



- Using the spanner supplied turn counternut **2** in a clockwise direction. Then turn counternut **2** together with foot **1** to unscrew

Figure 18: Overspecified reference using labels and functional information, ensuring that the user will be able to identify the functional element even if lacking technical knowledge.

In the case of the advanced washing machine, three different combinations of reference types were observed to yield overspecification. First, 20% out of the overspecified references combined repetition of visual elements in both the visual and text with references

providing semantic and/or functional information (see Figure 19). Second, 40% combined the repetition of words in both the visual and text with references providing perceptual information. Finally, 40% combined labeled elements in visuals with references providing semantic and/or functional information. While the last type was used with line drawings, the other two referred to visuals showing display content.

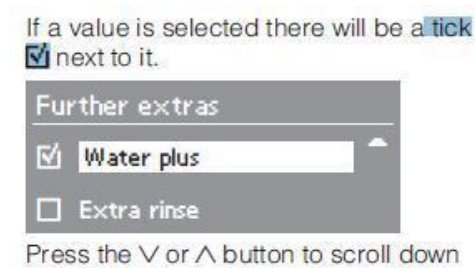


Figure 19: Example for an overspecific reference using visual elements and semantic information, drawing the reader's attention effectively to the visual element in the display.

CONCLUSIONS, LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this section, we review our results with respect to implications to practice and theory, and consider limitations and further steps needed to extend the present findings. This section starts with conclusions, followed by limitations, and suggestions for future research.

Conclusions

Our qualitative analysis of a small collection of manuals highlights a number of important considerations. This section starts with implications to practice followed by implications to research and theory.

Implications to Practice

The results of our study suggest that technical writers need to attend to the links between text and visuals carefully so as to support the user effectively. Purposeful strategies of direct reference and overspecification will help to avoid ambiguity and failed communication.

In particular, we found that the most frequently used type of reference – references providing semantic and/or functional information – can be ambiguous. While many other types of references use perceptual elements such as labels, arrows, or crosses to direct the

reader's attention to specific parts of the visual, the use of semantic and/or functional information as a reference relies on the reader's prior knowledge or understanding of the appliance. This may work well for cases where the relevant technical knowledge can reasonably be expected; however, users consulting a manual evidently feel a need for technical support – which may mean that they lack precisely the kind of knowledge that is needed here. In the case of our data set, the features of the associated visuals reduced the chance for misinterpretation. Nevertheless, this type of reference may be well be felt to be too imprecise for users with little or no prior knowledge. These issues need to be carefully considered by technical writers and designers.

Another issue to consider concerns the potential ambiguity in dimensional expressions such as *left*, *right*, *in front of*, and *behind*. In the case of our data set, these terms were rarely used, which may mean that the technical writers were aware of the possible misinterpretation. References to the vertical axis (*above*, *below*) are less problematic as they usually rely unambiguously on gravity.

Implications to Research and Theory

To aid practice in this field, research needs to establish more precisely which kinds of reference types provide most effective and unambiguous links for users. Although spatial proximity plays an important role in suggesting the relation between text and visuals, this is typically insufficient to establish the role of the visual information unambiguously for the user. The text that is used to refer to a visual does not only serve as a "pointer" to the accompanying illustration, but typically adds vital information to its message. Most often, this concerns semantic and functional information about specific elements shown in the visual. Other references inform about actions to be performed (represented textually by an action verb) as illustrated through arrows in the visual, and yet others point to and explain specific labels of elements in the visuals. Another frequently observed form of reference in the analyzed operating instructions is the embedding of visuals in the text. Here the reference

becomes part of the sentence, and considering the visual is essential to the understanding of the text. This range of options is not exhaustive: other types were found in our small data set, and it can be expected that enlarging the data would lead to the discovery of further types.

Crucially, together these options serve to establish meaningful and informative relations between text and visuals, directing the users' attention toward the relevant aspects and adding elaboration where necessary. In some cases, reference types were found to be combined, yielding overspecification in reference; this leads to cognitive supportive redundancy by guiding the user's attention in more than one way. Here, future research needs to address the efficiency of each of these types of links, so as to establish its user friendliness and cognitive effects on the readers.

Moreover, there are important differences between types and significance of visuals. The visuals in the operating instructions could be either essential or (to some extent) redundant – and this distinction might be quite essential for establishing efficient communication. While some visuals in our data set were essential to the understanding of the situation and the accompanying text (such as visuals that were embedded directly in the text), others offered additional information about the elements mentioned in the text but provided no essential and no completely new information. Many of the visuals that were referred to by semantic and functional information fall into this category, as well as those illustrating actions, among others. Furthermore, the number of references made to visuals appears to depend on the type of visual, as there were significantly more references to line drawings than to visuals showing display content in our data set. Also, while all of the visuals to which no references could be detected were visuals showing display content, line drawings were more likely to be overspecified. Since visuals can take many different forms according to the type of device described in a manual, it is to be expected that reference relations will also differ in this respect. This opens up a range of interesting research questions, such as: What are the reasons for different referencing strategies in writers of manuals – that is, what makes

technical writers believe that a certain type of visual is more self explanatory than another?

And to what extent is this true? Do users need different levels of referential information depending on the type of visual? Also, which kind of referential strategy is most efficient for which type of visual?

Limitations

Our research addressed the types and potential ambiguity of linking strategies between text and visuals on the basis of an in-depth investigation of only a small data set. Although we conducted an informal check of other manuals to address the validity and generalizability of our results across appliances, the study is clearly limited by the size of the data set investigated. Our study could therefore be viewed as an in-depth preliminary step preparing a more extensive investigation of linking strategies based on a balanced and representative selection of manuals; this would allow for insights on distributional patterns that our study cannot meaningfully provide. Furthermore, based on the available literature we suggested potential effects on the reader by pointing to potential ambiguities and the effects of overspecification, but did not include user studies to test these effects directly.

Suggestions for Future Research

Future research should address the effects of the various identified phenomena on readers and users of the operating instructions in practical applications, so as to identify differences in quality of user manuals. In particular, some types of visuals and reference relations may direct the reader's attention to relevant items more effectively than others; this should influence the cognitive processing of information as well as practical usability and user satisfaction.

REFERENCES

- [1] M.A.K. Halliday, *An Introduction to Functional Grammar*. London: Edward Arnold, 1994 (2nd Ed).
- [2] T. van Leeuwen, "Ten reasons why linguists should pay attention to visual communication," in *Discourse and Technology – Multimodal Discourse Analysis*, P.

LeVine and R. Scollon, Eds. Washington, D.C.: Georgetown University Press, 2004, pp. 7-19.

- [3] J.R. Martin and D. Rose, *Working with Discourse: Meaning Beyond the Clause*. Continuum International Publishing Group Ltd, 2003.
- [4] T. Tenbrink, "Cognitive Discourse Analysis: Accessing cognitive representations and processes through language data", *Language and Cognition*, vol. 7, no. 1, 98 – 137, 2015.
- [5] H. Leckie-Tarry (ed. by David Birch), *Language and Context: A functional linguistic theory of register*, London, New York: Pinter, 1995.
- [6] V. K. Bhatia, *Worlds of Written Discourse: A Genre-Based View*, London, UK: A&C Black, 2004.
- [7] P. Baggett, "Understanding visual and verbal messages," in *Knowledge Acquisition from Text and Pictures*, H. Mandl and J. R. Levin, Eds., Amsterdam, New York: Elsevier, 1989, pp.101-124.
- [8] G. Denhière and M. Denis. "The processing of texts describing spatial configurations," in *Knowledge Acquisition from Text and Pictures*, H. Mandl and J. R. Levin, Eds., Amsterdam, New York: Elsevier, 1989, pp.249-261.
- [9] P. N. Johnson-Laird, *Mental Models*, Cambridge, MA: Harvard University Press, 1983.
- [10] A. M. Glenberg and W. E. Langston, "Comprehension of illustrated text: Pictures help to build mental models," *Journal of Memory and Language*, vol. 31, pp. 129-151, 1992.
- [11] F. Ganier, "Factors affecting the processing of procedural instructions: implications for document design," *IEEE Transactions on Professional Communication*, vol. 47, no. 1, pp. 15-26, 2004.
- [12] J. Heiser, M. A. Doantam Phan, B. Tversky, and P. Hanrahan, "Identification and validation of cognitive design principles for automated generation of assembly instructions," in *Proceedings of Advanced Visual Interfaces*, 2004, pp. 311-319.

- [13] A. A. Maes and H. Lenting, "How to put the instructive space into words," *IEEE Transactions on Professional Communication*, vol. 42, no. 2, pp. 100-113, 1999.
- [14] M. Hegarty and M. A. Just, "Understanding machines from text and diagrams," in *Knowledge Acquisition from Text and Pictures*, H. Mandl and J. R. Levin, Eds., Amsterdam, New York: Elsevier, 1989, pp. 171-194.
- [15] A. Arts, A. Maes, L. Noordman, and C. Jansen, "Overspecification facilitates object identification," *Journal of Pragmatics*, vol. 43, pp. 361-374, 2011.
- [16] C. D. Rude and A. Eaton. Technical Editing. Longman, 2011.
- [17] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, pp. 1-39, 1986.
- [18] D. McNeill (Ed.) *Language and Gesture*. Cambridge, UK: Cambridge University Press, 2000.
- [19] K. Emmorey and S. Casey, "Gesture, thought, and spatial language," in *Spatial Language: Cognitive and Computational Aspects*, K. R. Coventry and P. Olivier, Eds. Dordrecht, Netherlands: Kluwer, 2002, pp. 87-101.
- [20] H. H. Clark, *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
- [21] A. Sloman, "Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence," *Artificial Intelligence*, vol. 2, no. 3-4, pp 209-225, 1971.
- [22] M. Gattis (ed.). *Spatial Schemas and Abstract Thought*. Cambridge, MA: MIT Press, 2003.
- [23] B. Tversky, "Some ways that maps and graphs communicate," in *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, C. Freksa, W. Brauer, C. Habel and K. F. Wender, Eds. New York, NY: Springer, 2000, pp. 72-79.

- [24] You, H., Young, R. and Lee, M. W. (2008). Development and usability testing of an operating manual for a hair-braiding machine. *Human Factors and Ergonomics in Manufacturing & Service Industries 18*: 1–13.
- [25] A. Herskovits, "Schematization," in *Representation and Processing of Spatial Expressions*, P. Olivier and K.-P. Gapp, Eds. Mahwah, NJ: Lawrence Erlbaum, 1998, pp. 149-162.
- [26] L. Talmy, "The fundamental system of spatial schemas in language," in *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, B. Hampe, Ed. Berlin, Germany: Mouton de Gruyter, 2005, pp. 199-234.
- [27] G. Kress and T. Van Leeuwen, *Multimodal Discourse. The modes and media of contemporary communication*, London, UK: Arnold Publishers, 2001.
- [28] R. Krull, S.J. D'Souza, D. Roy, and D.M. Sharp. "Designing procedural illustrations," *IEEE Transactions on Professional Communication*, vol. 47, no. 1, pp. 27–33, 2004.
- [29] A. Paivio, *Mental representations: A dual coding approach*, Oxford, UK: Oxford University Press, 1986.
- [30] S.-P. Ballstaedt (2005). *Visualisierung: Bilder in der technischen Dokumentation*. FH Gelsenkirchen. Gelsenkirchen, Germany. [lecture notes on Zertifikatslehrgang Technical Writing/Technische Dokumentation] Available: www.fh-gelsenkirchen.de/fb02/homepages/ballstaedt/inhalte/publikationen/publikationen.html
- [31] W. Winn, "The design and use of instructional graphics," in: *Knowledge Acquisition from Text and Pictures*, H. Mandl and J.R. Levin, Eds., Amsterdam, New York: Elsevier, 1989, pp. 125-144.
- [32] M. Gellevis, H. van der Meij, T. De Jong and J. Pieters, "The effects of screen captures in manuals: A textual and two visual manuals compared," *IEEE Transactions on Professional Communication*, vol. 42, no. 2, pp.77-91, 1999.

- [33] M. Florax and R. Ploetzner, "What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity," *Learning and Instruction*, vol. 20, no. 3, pp. 216-224, 2010.
- [34] A. Rothkegel, *Technikkommunikation [Technical Communication]*, Konstanz, Germany: UVK Verlagsgesellschaft mbH, 2010.
- [35] H. van der Meij, "Optimizing the joint handling of manual and screen," in *Minimalism Beyond the Nurnberg Funnel*, J. M. Carroll, Ed. Cambridge, MA, London, UK: MIT Press, 1998, pp. 275-310.
- [36] A. Arts, *Overspecification in instructive texts*. PhD dissertation, Tilburg University, Nijmegen, Netherlands: Wolf Publishers, 2004.
- [37] T. Tenbrink, *Space, time, and the use of language: An investigation of relationships*. Berlin: Mouton de Gruyter, 2007.
- [38] S. C. Levinson, *Space in Language and Cognition*. Cambridge University Press, 2003.
- [39] B. Tversky, S. Kugelmass and A. Winter, "Cross-cultural and developmental trends in graphic production," *Cognitive Psychology*, vol. 23, pp. 515-557, 1991.

Sources for operating instructions and illustrations

Gebrauchsanweisung Waschautomat W 1914 WPS de-DE (accessed via:

<http://miele.de/de/haushalt/produkte/manualshg.aspx>, last access: 14/05/2013)

Operating instructions for Washing machine W 1914 en-GB (accessed via:

<http://www.miele.co.uk/Owners/OperatingInstructions.aspx>

<http://www.miele.co.uk/Resources/OperatingInstructions/W%201914.pdf>, last access: 14/05/2013)

Gebrauchsanweisung Waschautomat W 5964 WPS de-DE (accessed via:

<http://miele.de/de/haushalt/produkte/manualshg.aspx>, last access: 14/05/2013)

Operating instructions for Washing machine W 5964 WPS en-GB (accessed via:

<http://www.miele.co.uk/Owners/OperatingInstructions.aspx>

<http://www.miele.co.uk/Resources/OperatingInstructions/W%205964%20wps.pdf>, last

access: 14/05/2013)